

SUPPORTING INFORMATION

Metabolomics Beyond Spectroscopic Databases: A Combined MS/NMR Strategy for the Identification of Metabolites in Complex Mixtures

Kerem Bingol,¹ Lei Bruschweiler-Li,² Cao Yu,² Arpad Somogyi,² Fengli Zhang,³ and Rafael Brüschweiler^{1,2,3}

¹Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States

²Campus Chemical Instrument Center, The Ohio State University, Columbus, Ohio 43210, United States

³National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310, United States

Contents

Figure S1: Direct infusion Q-TOF mass spectrum of ten-compound model mixture

Figure S2: Chemical structures of best hits of ten-compound model mixture

Figure S3: Direct infusion Q-TOF mass spectrum of *E. coli* cell lysate

Figure S4: 2D ¹³C-¹H HSQC spectrum of *E. coli* cell lysate

Figure S5: Chemical structures of best hits of *E. coli* cell lysate

Table S1: Molecular formulas obtained from direct infusion Q-TOF mass spectrum of ten-compound model mixture along with number of possible structures obtained with ChemSpider

Table S2: Experimental ¹³C-¹H HSQC cross-peak lists of nine metabolites in the model mixture

Table S3: Molecular formulas obtained from direct infusion mass spectrum of *E. coli* cell lysate along with the number of possible structures in ChemSpider database

Table S4: Experimental ¹³C-¹H HSQC peak lists extracted from 2D ¹³C-¹H HSQC spectrum of *E. coli* cell lysate

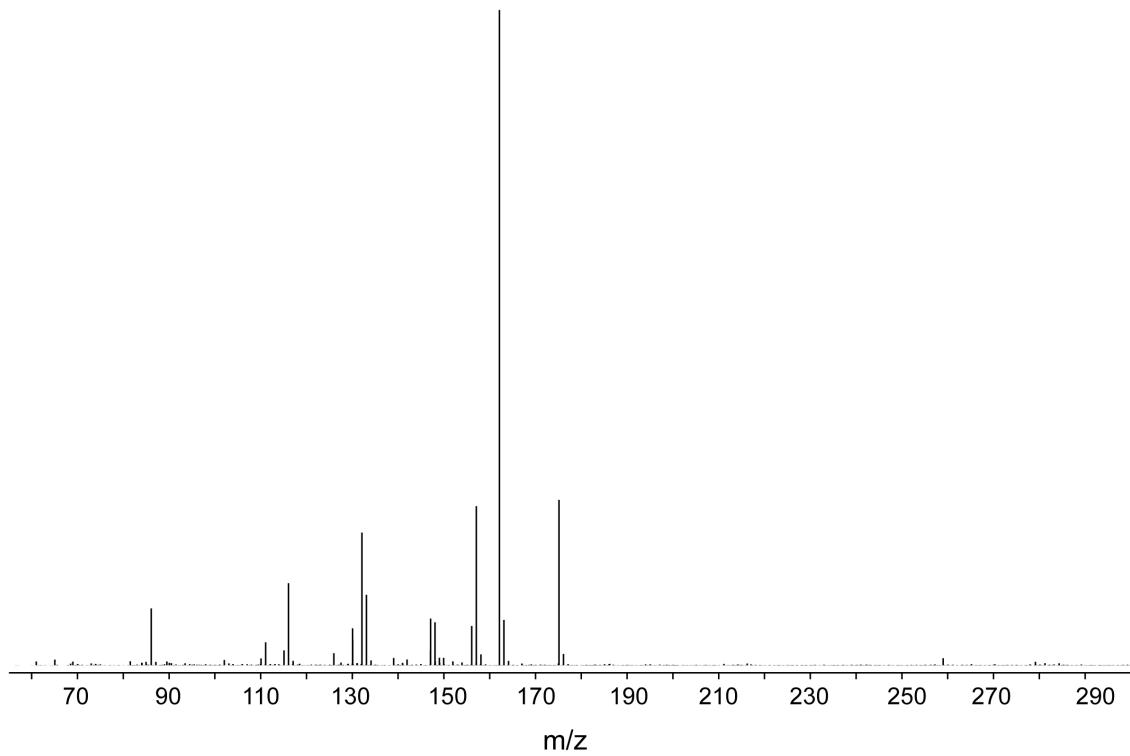


Figure S1. Direct infusion Q-TOF mass spectrum of ten-compound model mixture acquired in positive ion mode with a mass accuracy measurement error of $<\pm 5$ ppm. The mixture components are lysine, shikimate, carnitine, isoleucine, glutamate, histidine, arginine, alanine, ornithine, and glutamine.

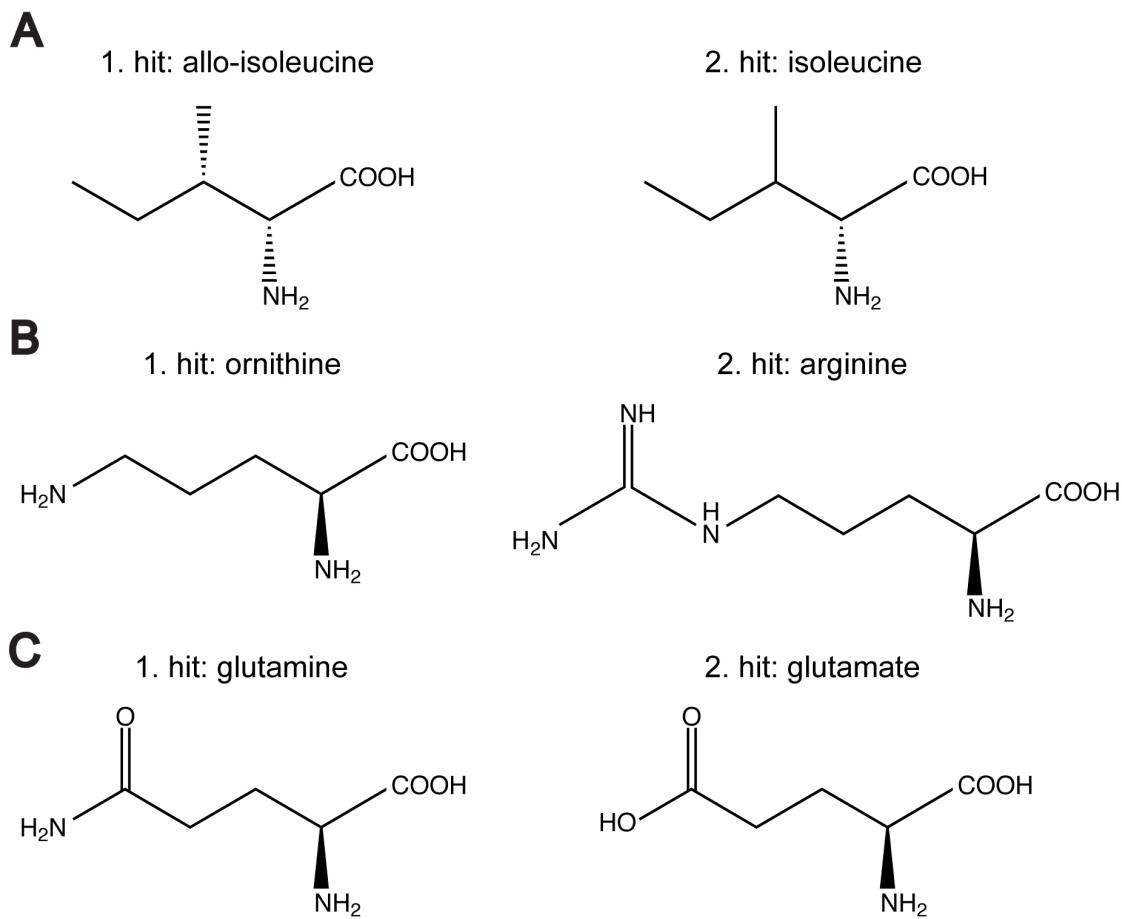


Figure S2. Chemical structures of top hits and second best hits when experimental 2D ^{13}C - ^1H spectra of (A) isoleucine, (B) arginine, and (C) glutamate contained in ten-compound model mixture were quantitatively compared with predicted ^{13}C - ^1H HSQC spectra of 4772 structures. Molecules returned as top hits (false-positives) are allo-isoleucine, ornithine, and glutamine. Their structures are notably similar to the structures of the true positives (i.e., isoleucine, arginine, and glutamate).

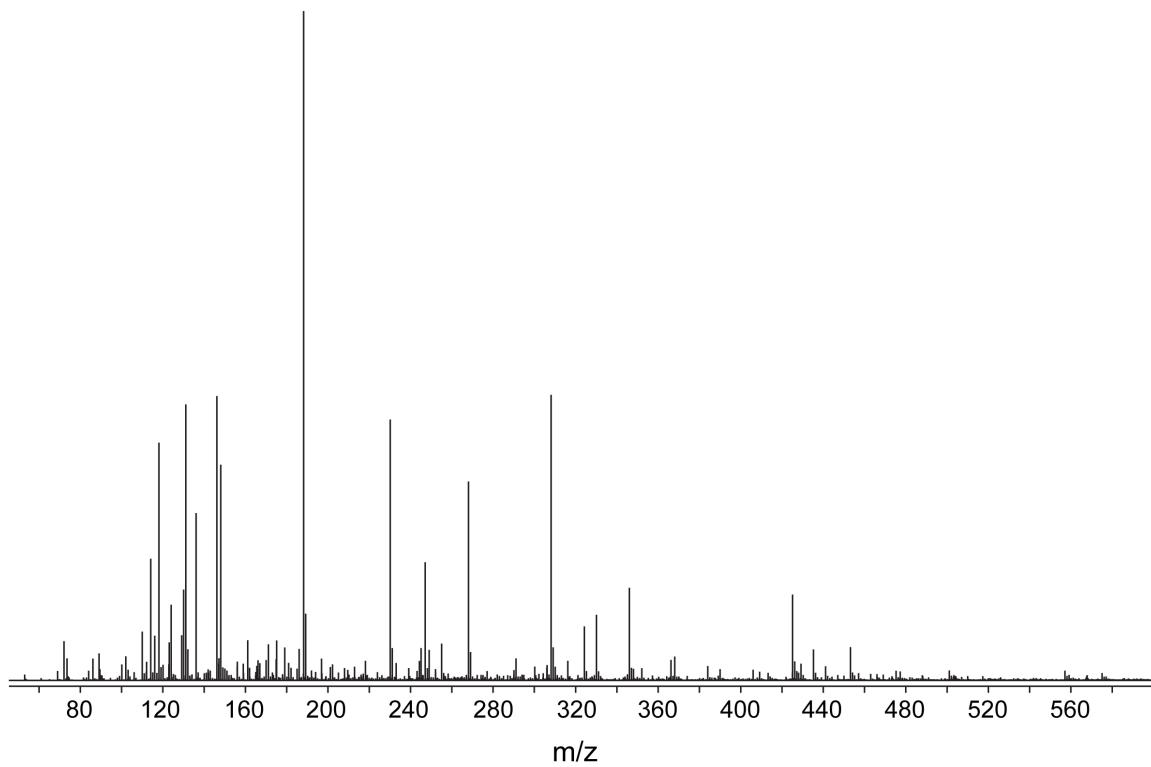


Figure S3. Direct infusion Q-TOF mass spectrum of *E. coli* cell lysate acquired in positive mode. The metabolites identified in the spectrum by using SUMMIT MS/NMR are listed in Table 2.

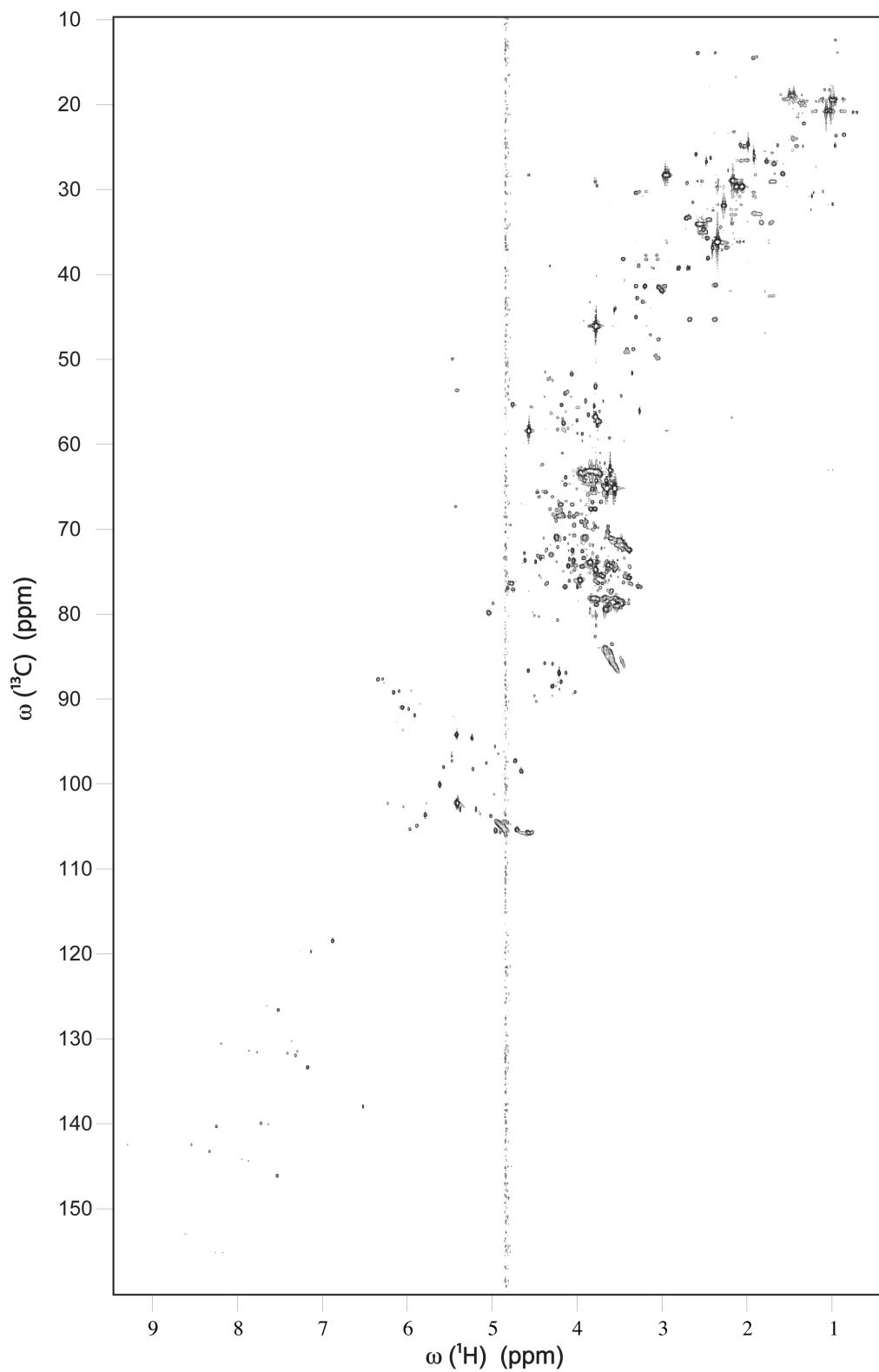


Figure S4. 2D ^{13}C - ^1H HSQC spectrum of *E. coli* cell lysate. The spectrum was acquired at 800 MHz proton frequency at 298 K with 512 and 1024 complex data points along the indirect and direct dimensions, respectively.

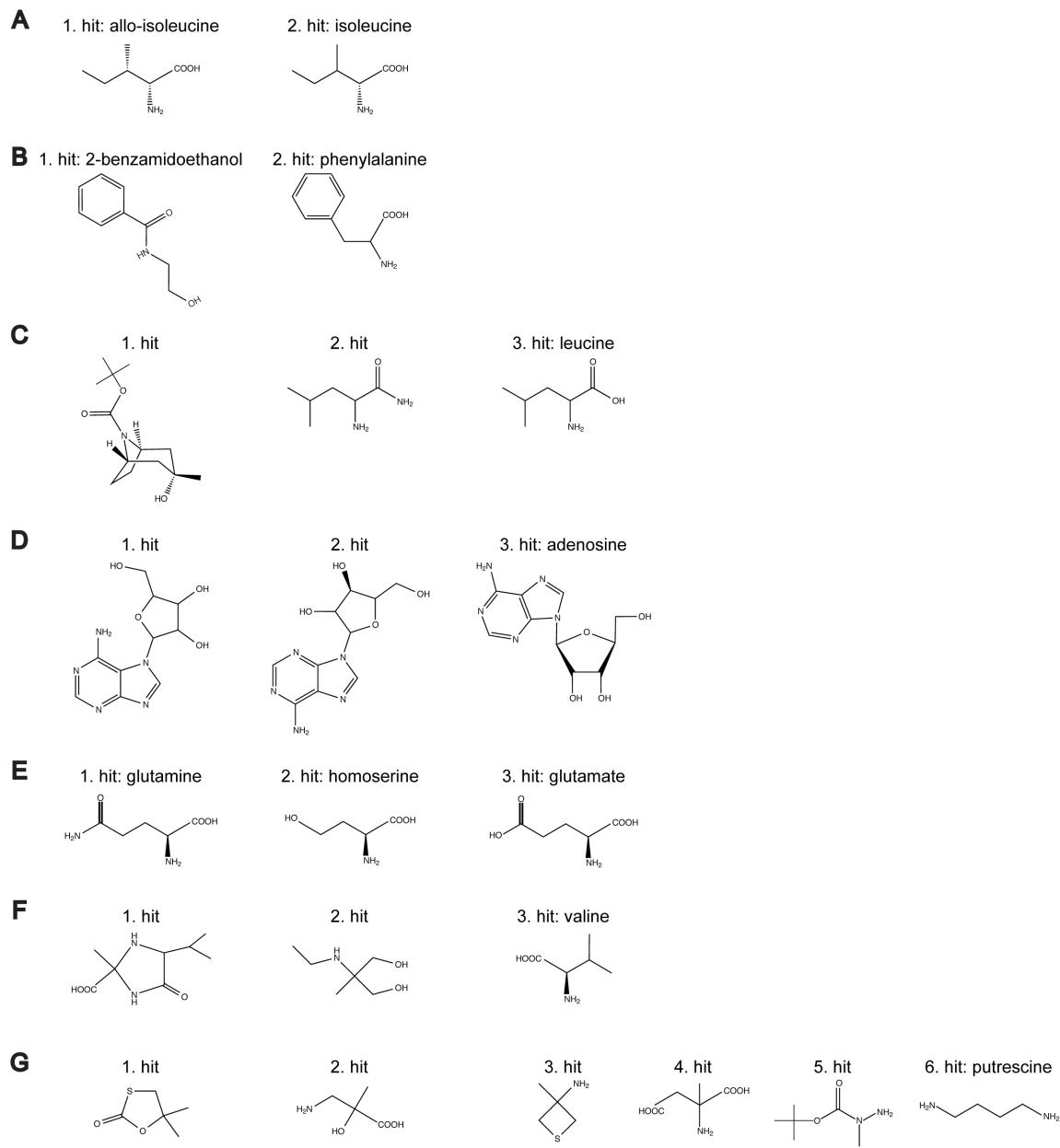


Figure S5. Best hit structures when experimental 2D ^{13}C - ^1H HSQC peaklist of (A) isoleucine, (B) phenylalanine, (C) leucine, (D) adenosine, (E) glutamate, (F) valine, and (G) putrescine from *E. coli* cell lysate were compared with predicted ^{13}C - ^1H HSQC spectra of 13,872 structures consistent with MS data.

Table S1. Molecular formulas obtained from direct infusion Q-TOF mass spectrum of ten-compound model mixture along with number of possible structures obtained with ChemSpider. From each of the 4772 structures, which constitute the total structural manifold, a ^{13}C - ^1H NMR HSQC spectrum was predicted for comparison with experimental NMR spectra.

Chemical formula	# of structures	Chemical formula	# of structures
C ₇ H ₁₅ NO ₃	362	C ₆ H ₁₁ NO ₂	678
C ₆ H ₁₄ N ₄ O ₂	65	C ₅ H ₁₀ N ₂ O	163
C ₄ H ₈ N ₆ O	25	C ₅ H ₁₀ N ₂ O ₃	176
C ₆ H ₁₃ NO ₂	535	C ₇ H ₁₇ N ₃ S	24
C ₅ H ₉ NO ₂	397	C ₇ H ₁₃ N ₂ S	274
C ₅ H ₁₂ N ₂ O ₂	241	C ₄ H ₇ NO ₂	195
C ₅ H ₁₁ N	112	C ₇ H ₁₇ NS	31
C ₆ H ₁₄ N ₂ O ₂	294	C ₄ H ₅ NO	72
C ₅ H ₉ NO ₄	167	C ₃ H ₇ NO ₂	74
C ₆ H ₉ N ₃ O ₂	514	C ₆ H ₁₀ OS	79
C ₅ H ₇ NO ₃	215	C ₅ H ₉ N	79

Table S2. Experimental ^{13}C - ^1H HSQC peak lists of nine metabolites in the model mixture. Each peak list was compared with the predicted ^{13}C - ^1H HSQC spectrum of each of the 4772 structures of the total manifold (Table 1).

Metabolite	^1H ppm ^{13}C ppm	Metabolite	^1H ppm ^{13}C ppm
Lysine	3.0116 41.7550 1.7163 29.1370 1.4647 24.1450 1.8953 32.6280 3.7453 57.1900	Arginine	3.7638 57.0030 1.9090 30.2810 1.6450 26.5760 1.7140 26.5760 3.2364 43.2000
Carnitine	3.2147 56.7950 3.4185 72.8170 4.5548 66.7850 2.4244 45.6820	Alanine	1.4661 18.8680 3.7669 53.2160
Isoleucine	0.9269 13.8340 1.2490 27.1740 1.4584 27.1740 1.9710 38.6140 0.9986 17.4110 3.6576 62.2490	Ornithine	3.0423 41.5570 1.7971 25.4660 1.9350 30.1810 3.7704 56.8010
Glutamate	2.3391 36.1980 2.0811 29.7270 3.7475 57.3570	Glutamine	2.4403 33.5210 2.1253 28.9450 3.7599 56.8260
Histidine	7.8269 138.8720 7.0668 119.5740 3.1352 30.6960 3.2281 30.6960 3.9807 57.4360		

Table S3. Molecular formulas obtained from direct infusion mass spectrum of *E. coli* cell lysate along with the number of structures in ChemSpider database. From each of the 13,872 structures, which constitute the total structural manifold of the sample, a ^{13}C - ^1H NMR HSQC spectrum was predicted for comparison with experimental NMR spectra.

Chemical formula	# of structures	Chemical formula	# of structures
C ₉ H ₂₁ N ₃ O	88	C ₅ H ₁₄ P ₂	2
C ₇ H ₁₉ N ₃	31	C ₆ H ₁₁ N ₃ O	229
C ₆ H ₁₄ N ₂ O	383	C ₅ H ₁₂ N ₂ O	193
C ₁₁ H ₂₃ N ₃ O ₂	157	C ₆ H ₁₀ N ₂ O ₃	261
C ₅ H ₁₁ NO ₂	333	C ₆ H ₁₅ NO ₂	152
C ₅ H ₉ NO ₄	167	C ₅ H ₁₄ NO ₄	7
C ₆ H ₁₁ NO	454	C ₆ H ₁₀ OS	79
C ₅ H ₇ NO ₃	215	C ₃ H ₇ NO ₂	75
C ₇ H ₁₆ N ₂	367	C ₄ H ₇ NO ₄	56
C ₅ H ₉ NO ₂	397	C ₅ H ₈ O ₂ S	88
C ₆ H ₁₂ N ₂ O ₃	274	C ₉ H ₁₇ NO ₃	731
C ₆ H ₁₄ N ₄ O ₂	65	C ₄ H ₉ NS	41
C ₉ H ₁₈ N ₂ O	718	C ₁₁ H ₂₀ N ₂ O ₃	581
C ₁₂ H ₂₆ N ₂ S	29	C ₁₂ H ₂₅ NO ₃	128
C ₆ H ₁₃ NO ₂	535	C ₇ H ₁₅ NOS	84
C ₄ H ₁₂ N ₂	40	C ₈ H ₁₄ N ₂ O ₃	312
C ₄ H ₇ NO ₂	195	C ₉ H ₂₁ NO ₄	11
C ₅ H ₁₁ N	112	C ₇ H ₂₀ N ₇	1
C ₇ H ₁₃ N	252	C ₆ H ₁₅ NO	248
C ₅ H ₁₀ N ₂ O ₃	176	C ₄ H ₁₃ P ₂	5
C ₅ H ₉ NO	223	C ₁₀ H ₁₃ N ₅ O ₄	340
C ₆ H ₁₄ S	11	C ₄ H ₉ NO ₂	164
C ₁₀ H ₂₀ N ₂ O ₂	728	C ₅ H ₁₁ NO ₂ S	136
C ₆ H ₁₂ O ₂ S	147	C ₆ H ₅ NO ₂	89
C ₅ H ₁₄ N ₂	80	C ₉ H ₁₁ NO ₃	999
C ₄ H ₅ NO	74	C ₄ H ₄ N ₂ O ₂	97
C ₁₃ H ₂₄ NO ₃	914	C ₉ H ₁₁ NO ₂	1161
C ₆ H ₁₄ N ₂ O ₂	295	C ₄ H ₉ NO ₃	142

Table S4. Experimental ^{13}C - ^1H HSQC peak lists extracted from 2D ^{13}C - ^1H HSQC spectrum of *E. coli* cell lysate. Each peak list was compared with the predicted ^{13}C - ^1H HSQC spectrum of each of the 13,872 structures of the total manifold (Table 2).

Metabolite	^1H ppm	^{13}C ppm	Metabolite	^1H ppm	^{13}C ppm
Aspartate	2.6880 2.8040 3.8990	39.2570 39.2570 54.8750	Putrescine	1.7700 3.0430	26.6520 41.4480
Alanine	1.4750 3.7830	18.9300 53.1870	Lysine	1.4410 1.7200 1.9020 3.0190 3.7470	23.9590 29.0400 32.7090 41.8720 57.4170
Betaine	3.2640 3.9040	56.0590 68.8300	Methionine	2.1310 2.1920 2.6380 3.8680	16.7330 32.4330 31.5150 56.5440
Adenosine	3.8400 3.9160 4.2970 4.4360 4.7880 6.0590 8.2170 8.3270	64.2100 64.2100 88.5000 73.2950 76.3770 90.9770 155.1880 143.2700	Nicotinic_acid	7.5180 8.2470 8.6030 8.9350	126.5920 140.3220 152.9950 151.5320
GABA	1.9170 2.2940 3.0090	26.6620 37.1060 42.1520	N_acetyl_putrescine	1.5790 1.6800 1.9850 3.0050 3.2010	28.1390 26.9570 24.6660 41.8320 41.3800
Glutamine	2.1520 2.4470 3.7830	29.0480 33.5480 56.8290	Spermidine	1.7730 1.7900 2.1070 3.0420 3.1020 3.1010 3.1370	26.7070 25.4760 26.7160 41.6140 39.2490 49.7030 47.1090
Arginine	1.6470 1.7170 1.9200 3.2290 3.7700	26.6450 26.6450 30.2850 43.2040 57.0090	Tyrosine	3.0510 3.1850 3.9400 6.8780 7.1730	38.2260 38.2260 58.7570 118.4620 133.3760
Isoleucine	0.9360 1.0130 1.2580 1.4640 1.9820 3.6730	13.8710 17.5860 27.2270 27.2270 38.5800 62.2180	Uracil	5.7880 7.5320	103.7810 146.1490
Leucine	0.9510 0.9620 1.7110 1.7120 3.7360	23.6680 24.8260 27.0630 42.5300 56.1060	Threonine	1.3250 3.5780 4.2540	22.2350 63.2370 68.7210
Phenylalanine	3.1270 3.2770 3.9860 7.3160 7.3700 7.4140	39.1910 39.1910 58.8550 132.0830 130.4160 131.7470	Valine	0.9910 1.0450 2.2790 3.6100	19.5750 20.9050 32.0220 63.2480
Glutamate	2.0990 2.3540 3.7680	29.6870 36.1700 57.2990			